

## Research Article

# Flower End-to-End Detection Based on YOLOv4 Using a Mobile Device

Zhibin Cheng<sup>1</sup> and Fuquan Zhang<sup>2</sup> 

<sup>1</sup>Educational Administration and Scientific Research, Fujian Polytechnic of Information Technology, Fuzhou, Fujian 350003, China

<sup>2</sup>College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China

Correspondence should be addressed to Fuquan Zhang; 8528750@qq.com

Received 30 July 2020; Revised 27 August 2020; Accepted 4 September 2020; Published 17 September 2020

Academic Editor: Chao-Yang Lee

Copyright © 2020 Zhibin Cheng and Fuquan Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, a novel flower detection application anchor-based method is proposed, which is combined with an attention mechanism to detect the flowers in a smart garden in AIoT more accurately and fast. While many researchers have paid much attention to the flower classification in existing studies, the issue of flower detection has been largely overlooked. The problem we have outlined deals largely with the study of a new design and application of flower detection. Firstly, a new end-to-end flower detection anchor-based method is inserted into the architecture of the network to make it more precious and fast and the loss function and attention mechanism are introduced into our model to suppress unimportant features. Secondly, our flower detection algorithms can be integrated into the mobile device. It is revealed that our flower detection method is very considerable through a series of investigations carried out. The detection accuracy of our method is similar to that of the state-of-the-art, and the detection speed is faster at the same time. It makes a major contribution to flower detection in computer vision.

## 1. Introduction

In recent years, flower classification and detection has been of considerable interest to the computer vision community which can be applied in AIoT for the smart garden. The flowers in the smart garden can be automatically designed and recommended to make it more beautiful. The previous work mostly focused on flower classification [1–5] using a traditional detector and method [6, 7]. While it has become a tendency in flower classification and detection based on deep learning anchor-based approaches, flower detection was paid little attention. In most studies of deep learning anchor-based flower detection, the approaches can be divided into two general classes, two-stage object detection [8–10] and one-stage object detection [11–14].

*1.1. Flower Classification and Detection Based on Deep Learning Two-Stage Approaches.* As we all know, flower detection has become a hot topic in object detection since the convolutional neural network reborns worldwide in 2012. And mainstream flower detection was divided into two

categories, one-stage detector approaches and two-stage detector approaches, which were based on anchor approaches. The essence of an anchor is the candidate boxes, which are designed with different scales and proportions and are classified by DNN. The positive anchor can learn how to return it to the right place, and it plays a role which is similar to the sliding window mechanism in traditional detection algorithms. In recent years, many researchers perform flower and fruit classification and detection based on convolutional neural network (CNN) approaches [15–18], which is a kind of feedforward neural network convolution computation contained and a deep structure. And it is one of the representative algorithms commonly used in deep learning. In 2014, Girshick et al. proposed a region with CNN characteristics (RCNN) of the first two-stage object detection (RCNN). And in 2015, they presented Fast RCNN, which enables us to train both the detector and the bounding box regressor in the same network configuration. Based on these models, in 2015, Ren et al. claimed a Faster RCNN detector, which is the first end-to-end and almost in a real-time deep learning detector (Faster RCNN) which proposal detection, feature extraction, bounding box

regression, and so on have been gradually integrated into a unified end-to-end learning framework in Faster RCNN. So many researchers have applied them in flower and all kinds of fruit detection including all kinds of fruit like mangoes, almonds, and apples [19–22].

Great progress has been made in flower detection based on two-stage approaches in high accuracy. However, it is concluded that the speed of it needs to be increased.

*1.2. Flower Classification and Detection Based on Deep Learning One-Stage Approaches.* To overcome the above limitations, considering the speed of the running time, the other one-stage flower detection method was demonstrated. Redmon et al. in 2015 demonstrated the YOLO one-stage detector, which divides the image into several regions and predicts the boundary box and probability of each region. It has fast detection speed which can contribute to processing streaming media video in real time. Compared with other algorithms such as the two-stage detection method, it has half as many false backgrounds as them on account of capturing contextual information effectively and a strong versatility and generalization ability. Then, Redmon et al. made a series of improvements based on YOLO [12], including YOLOv2 [13] and YOLOv3 [11], which has the same accuracy as the two-stage detector at a higher speed. It is faster in YOLOv2 than other detection systems in a variety of monitoring data sets, besides the tradeoff between speed and accuracy can be made. As we all know, improving recall and positioning accuracy was focused on in YOLOv2, while maintaining classification accuracy. However, in YOLOv3, on the premise of maintaining the speed advantage, the prediction accuracy is improved; particularly, the recognition ability of small objects is strengthened. It adjusted the network structure and multi-scale feature object detection method used and softmax utilized for object classification that counts. Therefore, many efforts were made to perform flower and fruit detection based on these algorithms of the YOLO detection method [23–25]. While its detection speed has been greatly improved compared with the two-stage detector, the positioning accuracy of it has been reduced, especially for some small objects. To overcome these limitations, Liu et al. in 2015 presented a Single Shot MultiBox Detector (SSD) to raise accuracy, especially small objects. The main contribution of SSD is the introduction of multireference and multiresolution detection technology to perform the detection only on its top layer. So many fruit and flower detection methods were based on SSD [26, 27], which achieved great success.

While great progress has been made in one-stage flower detection and has become a tendency in recent years, however, to obtain abundant image information and enhanced accuracy is still a challenge. Most of the studies were focused on accuracy or speed; however, few studies have considered the accuracy coming up with the speed of the object detection. So it is emphasized that the accuracy should be matched with the speed of the detector in our method. In conclusion, flower detection is based on one-stage approaches working well; the accuracy is still to be enhanced a little.

Although current anchor-based deep learning flower detection methods work well, they still suffer from the following six

problems: (1) Due to the irregular shape of the flowers, the bounding box covers a great deal of nonflower regions, which caused a lot of interference. (2) The setting of anchor needs to be designed manually, and different designs are required for different flower datasets, which are quite troublesome and does not conform to the design idea of DNN. (3) The matching mechanism of an anchor makes the frequency of extreme scale (very large and very small object) to match lower than that of moderate size object. It is not easy for DNN to learn these extreme flower samples well when learning. (4) The large number of anchors causes serious imbalance. (5) On account of the unlabelled flower dataset, which only can be trained in the flower classification model, it cannot perform flower detection in a mobile device and AIoT. (6) Labelling the flower dataset consumes a lot of time and power, therefore most of the researchers were not able to pay attention to the flower detection which can be applied in a mobile device and AIoT.

In order to overcome the above limitations, a new object detection method based on new anchor-based approaches and a new-labelled flower dataset is adopted in this paper. To acquire more useful information and more precious object position in the image, attention mechanism SAM is utilized, which the output feature map of the channel attention module is taken as the input feature map of this module. In our paper, we can regard SAM as an attention mechanism to be applied to both channel and spatial dimensions and it can be embedded in most of the current mainstream networks and can improve the feature extraction capability of network models without significantly increasing the amount of computation and parameters. The baseline of our dataset is Oxford 102 Flower dataset, and we have labelled the dataset with all categories and all label and geometry positions to perform flower detection and other plant detection in a smart garden. Our backbone network is the CSPDarnet53 network, and it was designed to solve the previous work in the reasoning process which requires a lot of computation from the point of view of network structure. Also, SSP block is added in CSPDarknet53 and it significantly increases the acceptance field, extracts the most important contextual characteristics, and does little to slow down network operations. Our contributions are summarized as follows:

- (i) We present a flower detection method, an end-to-end deep convolutional neural network for flower detection applied in a smart garden in AIoT. The backbone network we applied is the CSPDarnet53 network, which can reduce computation when ensuring accuracy by integrating gradient changes into feature maps from end to end. And the running time of the flower detection in the architecture we used is the state-of-the-art fastest model compared with other models, especially when integrated into the mobile device. In order to extract the most important contextual characteristics and also increase the acceptance field, SSP block is added in the backbone network when keeping the network operations. Besides, attention mechanism SAM is utilized to select the information that is more critical to the current task target from the numerous information

- (ii) We labelled Oxford 102 Flower dataset with annotation containing 102 categories of flowers common in the UK, with each category containing 40 to 258 images, for a total of 8,189 images to perform flower detection. As we all know, the previous work attached much importance to the flower classification with an unannotated dataset. In our work, flower detection was paid much attention to with the annotation flower dataset which can be trained in our architecture. Also, the flower detection we proposed can be integrated into the mobile device to make it convenient to operate in a smart garden which can be applied in flower arrangement and flower horticulture. What is more, it can not only enhance the user's human-computer interaction experience of flower arrangement but save time and cost of horticulturists. It has preliminarily realized the concept of the smart garden in AIoT.

The paper is organized as follows. Section 2 describes the materials and methods in our paper including the network architecture and the dataset we labelled. Section 3 describes our results and discussion. We conclude in Section 4.

## 2. Materials and Methods

The architecture of the flower detection system is shown as in Figure 1, when the flower picture is acquired by the mobile device which is based on 5G networks meaning that Internet of Everything; it is passed through our backbone network, the neck module with SPP and PAN and the module of YOLOv3 head. At last, the result of the flower picture can be obtained which can be applied in the smart garden in AIoT.

*2.1. Network Architecture.* The architecture of our network is shown in Table 1, and the backbone network we utilized is the CSPdarknet53 network, combining CSPNet [28] and darknet53. The CSPNet was designed to solve the previous work in the reasoning process which requires a lot of computation from the point of view of network structure. It is considered that the problem of high inference computation is caused by the gradient information repetition in network optimization while CSPNet can reduce computation when ensuring accuracy by integrating gradient changes into feature maps from end to end. It not only can enhance the learning ability of the CNN but also can reduce computing bottlenecks and memory costs while keeping accuracy in the lightweight. It is a utilized idea ResNET for reference and adding a residue module to the Darknet53 network which can help to solve the deep network gradient problem. Two convolutional layers and one shortcut connection are contained in each residue module, and there are several duplicate residual modules in the layers. The pool layer and full connection layer were not involved in the architecture; the network undersampling is achieved by setting the stride of convolution as 2. After passing through this convolutional layer, the size of the image will be reduced to half. Convolution, BN, and Leaky Relu are contained in each convolutional layer, and a zero padding is added after each residual module.

It is studied that CSPDarkNet53 has advantages in object detection, which is better as the backbone of the test model. In CSPDarknet53, the parameters of CSPDarknet53 for image classification are 27.6M which is bigger than other neural networks, and the receptive field size is also much bigger than neural networks. So it is a very suitable backbone for our proposed flower detection method.

In addition, the SSP block is attached to the backbone network CSPDarknet53, which can produce an output of fixed size regardless of input size and can use different dimensions of the same image as input to get pooling features of the same length. When SPP is placed behind the last convolutional layer, it has no influence on the structure of the network and just replaces the original pooling layer. It can be used not only for image classification but also for object detection. When the SSP block is made use of in CSPDarknet53, it significantly increases the acceptance field, extracts the most important contextual characteristics, and does little to slow down network operations.

Also, PANET [29] was used instead of FPN in YOLOv3 as a parametric polymerization method for different bone levels for different detector levels. It is a bottom-up path enhancement that is aimed at facilitating the flow of information which can contribute to shortening the information path, enhancing the feature pyramid, and accurately locating the signal present at low levels to enhance the whole feature level. PANET developed adaptive feature pools to connect the feature grid to all the feature layers to enable useful information from each feature layer to propagate directly to the proposed subnetwork below.

YOLOv3 [11] is utilized in the head of the architecture, which is an anchor-based detection model. The residual network structure is used for reference to form a deeper network level and multiscale detection in the YOLOv3 model, which can improve mAP and small object detection effect.

*2.2. Flower Dataset.* As shown in Figure 2, Oxford 102 Flower dataset is a common flower dataset used in research and experiment when it comes to plant field. It contains 102 categories of flowers common in the UK, with each category containing 40 to 258 images, for a total of 8,189 images. Large proportions, gestures, and light changes are involved in the images which can be used for image classification studies. However, it cannot be used for flower detection when it comes to a smart garden in AIoT. Besides, the unlabelled flower dataset cannot be applied in a mobile device and other smart devices.

In our work, the Oxford 102 Flower dataset is labelled not only containing flower classification to perform flower detection to be utilized in a smart garden in AIoT. As shown in Figure 3, the labelled flower dataset can be contributed to all flower, plant, and fruit detection studies, especially in a completed natural environment. The dataset was labelled by a professional data annotator to classify the flower dataset which can contribute to the accuracy of the flower detection model and make it easier to perform real-time flower detection.

More training data can lead to a more sound model. If there are limited data volumes, it can make use of data augmentation to increase the diversity of the training sample to enhance model robustness, avoid overfitting, and improve

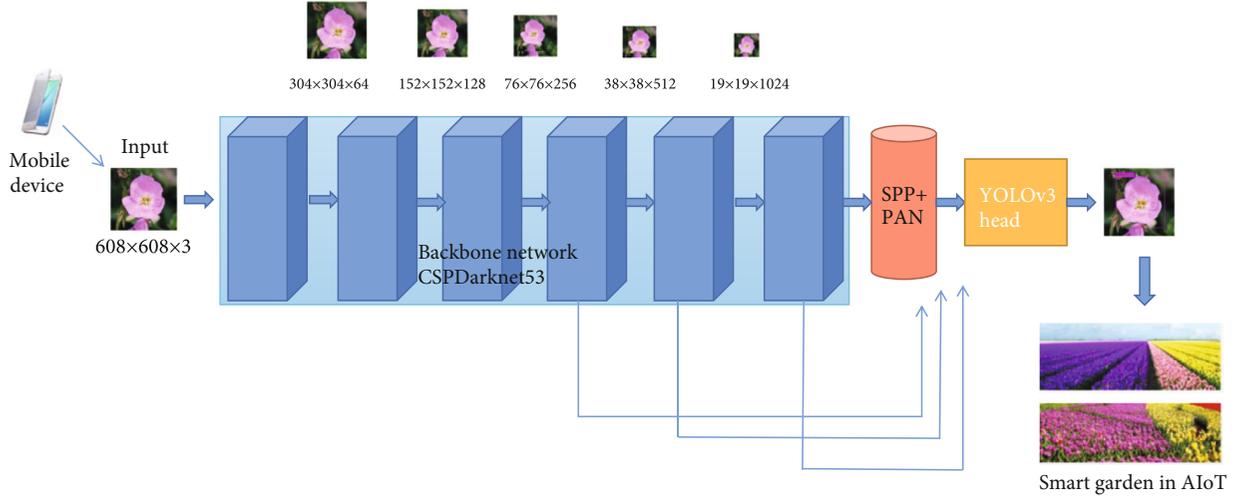


FIGURE 1: The architecture of our flower detection system.

TABLE 1: The architecture of the backbone network CSPDarknet53.

Type	Filters	Output
DarknetConv2D		
BN	32	$608 \times 608$
Mish		
ResBlock	64	$304 \times 304$
$2 \times$ ResBlock	128	$152 \times 152$
$8 \times$ ResBlock	256	$76 \times 76$
$8 \times$ ResBlock	512	$38 \times 38$
$4 \times$ ResBlock	1024	$19 \times 19$

the generalization ability of the model. The common methods include flip, rotation, shift, resize, random crop or pad, color jittering, and noise. In our dataset, data augmentation is also used to enhance our model.

**2.3. Loss Function.** In our work, the loss function DIoU loss is the loss function we used, which is feasible to directly minimize the normalized distance between the anchor frame and the target frame to achieve faster convergence speed and is more accurate and faster when overlapped or even included with the target box when making regression.

DIoU loss is on the basis of IoU (Intersection over Union), considering the information of the center distance of the bounding box. The definition of it can be defined as follows, formula (1), where  $B^{gt}$  is the target box and  $B$  is the prediction box. And the loss function of the IoU is defined as formula (2); it is shown that it works if the bounding boxes overlap; there is no overlap if the gradient does not change.

$$IoU = \frac{B \cap B^{gt}}{B \cup B^{gt}}, \quad (1)$$

$$L_{IoU} = 1 - \frac{B \cap B^{gt}}{B \cup B^{gt}}. \quad (2)$$

Therefore, it is GIoU that can improve the loss of IoU in the case that the gradient does not change without overlapping boundary boxes, which adds a penalty term on the basis of the loss function of IoU. It is defined as the following formula (3). The additional parameter  $C$  in the formula represents the minimum boundary box that can cover both  $B$  and  $B^{gt}$ . However, if one of  $B$  or  $B^{gt}$  overrides the other box in the case, the penalty term cannot work, which can be regarded as an IoU loss.

$$L_{GIoU} = 1 - IoU + \frac{|C - B \cap B^{gt}|}{|C|}. \quad (3)$$

To solve the limitations, the DIoU is introduced which can be defined as formula (4), where  $b$ ,  $b^{gt}$  represents the center point of the anchor frame and target frame, respectively, and  $p$  is the Euclidean distance between two centers. Also,  $c$  represents the diagonal distance of the minimum rectangle that can cover the anchor and target box simultaneously. What is more, the normalized distance between the anchor box and the target box is modeled in DIoU. The loss function of DIoU can be defined as formula (5).

$$R_{DIoU} = \frac{p^2(b, b^{gt})}{c^2}, \quad (4)$$

$$L_{DIoU} = 1 - IoU + \frac{p^2(b, b^{gt})}{c^2}. \quad (5)$$

Besides, CIoU is also proposed to our loss function. There are several advantages in the CIoU loss function: (1) It can increase the overlap area between the ground truth box and the prediction box. (2) It can minimize the distance between the center points. (3) It can keep the frame height ration consistent.

	Aphine sea holly	43		Buttercup	71		Fire lily	40
	Anthurium	105		Californian poppy	102		Foxglove	162
	Artichoke	78		Camelia	91		Frangipani	166
	Azalea	96		Canna lily	82		Fritillary	91
	Ball moss	46		Canterbury bells	40		Garden phlox	45
	Balloon flower	49		Cape flower	108		Gaura	67
	Barbeton daisy	127		Carnation	52		Gazania	78
	Bearded iris	54		Cautleya spicata	50		Geranium	114

FIGURE 2: Oxford 102 Flower dataset, which contains 102 categories of flowers, can be used for image classification studies.



FIGURE 3: Oxford 102 Flower dataset with annotation, which contains 102 categories of flowers, can be used for flower detection studies.

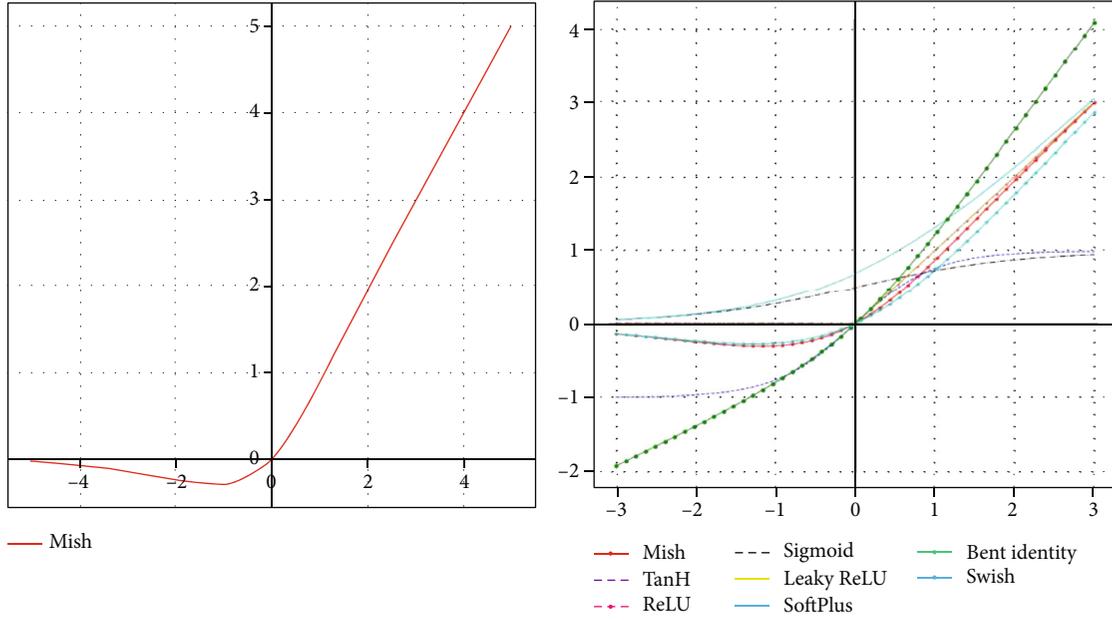


FIGURE 4: Representation of Mish activation function and its performance: (a) Mish activation function; (b) comparison of the different commonly used activation.

The CIoU can be defined as formula (6) based on formula (4), where  $\alpha$  represents a positive trade-off and  $v$  is defined in formula (7) to measure the consistency of the aspect ratio added.

$$R_{\text{CIoU}} = \frac{p^2(b, b^{\text{gt}})}{c^2} + \alpha v, \quad (6)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2. \quad (7)$$

The loss function of CIoU can be defined as formula (8), and the trade-off parameter  $\alpha$  can be defined in formula (9).

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{p^2(b, b^{\text{gt}})}{c^2} + \alpha v, \quad (8)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v}. \quad (9)$$

In our work, the DIoU loss function can be applied in NMS (Nonmaximum Suppression) to delete the redundant detection box. Not only the overlapping area is considered but also the distance between the detection box, and the center point of the target box is taken into account, which can effectively avoid the above two loss function mistakes.

**2.4. Activation Function.** Activation function is a function running on the neuron of the neural network, which is responsible for mapping the input of the neuron to the output. Its function is to increase the nonlinear change of the neural network model. In Figure 4, the Mish activation function and the comparison of different commonly used activation function are shown, in which the original figure can be found in Mish [30].

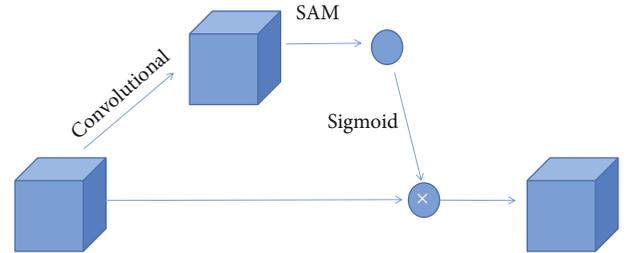


FIGURE 5: The modified SAM architecture.

In our architecture, the Mish [30] activation function is the activation function we utilized, which replaces Leaky Relu which is a very small constant leak that contained the improved function of Relu with Mish in YOLOv3 [11]. Leaky Relu is a self-regular nonmonotone neural activation function and a smooth activation function allowing better information into the neural network to obtain better accuracy and generalization. It can be defined as formula (10), where in the formula (10), it is shown that  $\zeta(x) = \ln(1 + e^x)$ .

$$f(x) = x \cdot \tanh(\zeta(x)). \quad (10)$$

It is more accurate and precious than Swish [31] defined as formula (12) and Relu [32] defined as formula (11) when performing on the experiments.

$$f(x) = \max(0, x), \quad (11)$$

$$f(x) = x \cdot \text{sigmoid}(x). \quad (12)$$

**2.5. Attention Mechanism.** As we all know, attention mechanism has become an important part of the structure of neural



FIGURE 6: The results of flower detection.

networks and has a large number of applications in fields such as natural language processing, statistical learning, and computer vision in the field of artificial intelligence. The attention mechanism in deep learning is similar to the selective visual attention mechanism of human beings in essence. The core goal is to select the information that is more critical to the current task target from the numerous information.

The SE [33] and SAM [34] model is the common attention mechanism used in deep learning networks. The purpose of the SE model is to reweight to the feature channel and only pay attention to which layer on the channel level will have stronger feedback ability, but it cannot reflect the meaning of “attention” on the spatial dimension, while SAM which is the output of the feature map of the channel attention module that is taken as the input feature map of this module is more adaptive to our architecture since it saves a lot of computing resources. In our work, the modified SAM has been utilized, which is shown in Figure 5. It is seen that spacewise attention is used instead of pointwise attention in it. Besides, PAN is replaced with concatenation connections.

### 3. Results and Discussion

We describe the results of our flower detection in Figure 6, which shows the result of flower detection containing category and degree of confidence added in each category. The confidence levels are 98%, 98%, 94%, and 84%, respectively,

in the test results, which achieved the desired effect. It is demonstrated that the flower detection can be achieved at a high standard which the accuracy has matched the speed. The GPU we used are 2 Titan Xp. The basic requirements for a camera on a mobile device are based on HUAWEI P20 Pro, and the processor is HiSilicon Kirin 970. When applied in a smart garden in AIoT, the basic mobile network relied on a 5G network, which is a high-speed, low-delay, low-power consumption and ubiquitous network.

### 4. Conclusions

In conclusion, it seems that flower detection based on the state-of-the-art method is very considerable. Although widely accepted, it suffers from some limitations due to the flower detection algorithms that have not been integrated into the mobile device completely to perform the final application results of our flower detection to enhance user’s human-computer interaction experience in a smart garden combining with virtual reality. It is a tendency for the flower detection proposed in our paper combined with virtual reality. In a smart garden, it can be realized that flowers can be arranged intelligently not depending on people just on the smart phone or other mobile devices. Furthermore, intelligent gardening can come true which can reduce the cost of human and finance and it just relies on the human-

computer interaction. In the future work, we will pay much importance to it to accomplish the goal in our future work.

### Data Availability

The data we used is available, which is named Oxford 102 Flowers that part of them can be accessed and downloaded to perform the flower classification research and other experiments. And in our work, our dataset was labelled based on this dataset with annotation. So part of them with annotation is available to you from the corresponding author upon request (8528750@qq.com).

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the study of this work and publication of this paper.

### Acknowledgments

Our work is supported by the National Natural Science Foundation of China, Project No.: 61503082, Project name: Research on Large-Scale Ontology Matching Based on Evolutionary Algorithm.

### References

- [1] A. D. Aggelopoulou, D. Bochtis, S. Fountas, K. C. Swain, T. A. Gemtos, and G. D. Nanos, "Yield prediction in apple orchards based on image processing," *Precision Agriculture*, vol. 12, no. 3, pp. 448–456, 2011.
- [2] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007.
- [3] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, June 2006.
- [4] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [5] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pp. 1447–1454, New York, NY, USA, June 2006.
- [6] J. Lu, W. S. Lee, H. Gan, and X. Hu, "Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis," *Biosystems Engineering*, vol. 171, pp. 78–90, 2018.
- [7] L. Haozhou, C. Lipeng, M. Longtao, G. Zongbin, and C. Yongjie, "A recognition method of kiwifruit flowers based on K-means clustering," *Journal of Agricultural Mechanization Research*, vol. 42, no. 2, pp. 22–26, 2020.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017.
- [9] R. Girshick, "Fast r-cnn," *2015 IEEE International Conference on Computer Vision (ICCV) IEEE*, 2016.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2014.
- [11] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, <http://arxiv.org/abs/1804.02767>.
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [13] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.
- [14] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multi-box detector," *Computer Vision – ECCV 2016*, 2016.
- [15] P. A. Dias, A. Tabb, and H. Medeiros, "Multispecies fruit flower detection using a refined semantic segmentation network," *IEEE Robotics & Automation Letters*, vol. 3, no. 4, pp. 3003–3010, 2018.
- [16] B. Shree, R. B. Divya, and N. S. Rani, "Fruit detection from images and displaying its nutrition value using deep Alex network," in *Soft Computing and Signal Processing. Advances in Intelligent Systems and Computing*, vol. 898, J. Wang, G. Reddy, V. Prasad, and V. Reddy, Eds., Springer, Singapore, 2019.
- [17] K. Bresilla, G. D. Perulli, A. Boini, B. Morandi, L. C. Grappadelli, and L. Manfrini, "Comparing deep-learning networks for apple fruit detection to classical hard-coded algorithms," *Acta Horticulturae*, vol. 1279, no. 1279, pp. 209–216, 2020.
- [18] S. Bargoti and J. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *Journal of Field Robotics*, vol. 34, no. 6, pp. 1039–1060, 2017.
- [19] S. Wan and S. Goudos, "Faster R-CNN for multi-class fruit detection using a robotic vision system," *Computer Networks*, vol. 168, article 107036, 2020.
- [20] J. Lim, H. S. Ahn, M. Nejati, J. Bell, H. Williams, and M. D. BA, *Deep neural network based real-time kiwi fruit flower detection in an orchard environment*, 2020.
- [21] P. Lin, W. S. Lee, Y. M. Chen, N. Peres, and C. Fraisse, "A deep-level region-based visual representation architecture for detecting strawberry flowers in an outdoor field," *Precision Agriculture*, vol. 21, no. 2, pp. 387–402, 2020.
- [22] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–8, Singapore, Singapore, June 2017.
- [23] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417–426, 2019.
- [24] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'," *Precision Agriculture*, vol. 20, no. 6, pp. 1107–1135, 2019.
- [25] J. Zhao and J. Qu, "A detection method for tomato fruit common physiological diseases based on YOLOv2," in *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, Qingdao, China, China, August 2019.

- [26] H. X. Peng, B. Huang, Y. Y. Shao et al., “General improved SSD model for picking object recognition of multiple fruits in natural environment,” *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 16, pp. 155–162, 2018.
- [27] Q. Liang, W. Zhu, J. Long, Y. Wang, W. Sun, and W. Wu, “A real-time detection framework for on-tree mango based on SSD network,” in *International Conference on Intelligent Robotics & Applications*, Springer, Cham, 2018.
- [28] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, “CSPNet: A New Backbone that can Enhance Learning Capability of CNN,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop)*, Seattle, WA, USA, USA, June 2020.
- [29] S. Liu, Q. Lu, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8759–8768, Salt Lake City, UT, USA, June 2018.
- [30] D. Misra, “Mish: a self regularized non-monotonic neural activation function,” 2019, <http://arxiv.org/abs/1908.08681v2>.
- [31] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” 2017, <http://arxiv.org/abs/1710.05941>.
- [32] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” *Proceedings of International Conference on Machine Learning (ICML)*, pp. 807–814, 2010.
- [33] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.
- [34] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: convolutional block attention module,” *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.